

模型传递用于解析遥感傅里叶变换红外谱图

张琳, 张黎明, 李燕*, 胡兰萍, 王俊德

南京理工大学现代光谱研究室, 江苏南京 210014

摘要 基于模型传递的原理, 建立了一种可对存在背景干扰, 谱峰严重混叠的遥感傅里叶变换红外(remote sensing Fourier transform infrared; RS-FTIR)谱图进行解析的方法。分别用4组分气体混合物的EPA红外标准谱图和RS-FTIR谱图, 建立校正和预测模型, 经过正交信号校正(OSC)处理后, 用EPA红外谱图数据所建立的校正模型, 对RS-FTIR谱图数据进行预测, 得到的丙酮、甲醇、苯和三氯甲烷的均方根预测误差(RMSEP)分别为: 0.008 5, 0.018 0, 0.064 0, 0.002 8。未经OSC处理时的RMSEP依次是0.085 6, 0.047 9, 1.065 3, 0.014 2。经优化, 支集选择的方法为Kennard-Stone法, OSC在实现过程中循环次数为3时得到的预测结果最好。研究结果表明, 该方法能够克服背景和校正模型给RS-FTIR监测大气污染物带来的制约。

主题词 正交信号校正; 模型传递; 谱图解析

中图分类号: O657.3 **文献标识码:** A **文章编号:** 1000-0593(2006)09-1614-04

引言

由于具有灵敏、专属、快速等优势, 遥感傅里叶变换红外光谱(RS-FTIR)在空气污染物监测方面受到越来越多的重视^[1, 2]。但是, 两大问题限制了该技术的广泛应用^[3], 一是由于环境中温度、湿度等变化导致收集到的背景不稳定, 而造成吸收谱图常有基线漂移等问题, 为此, 出现了各种数据预处理方法, 如: 二级导数(SOD)^[4]和多倍分散校正(MSC)^[4]等; 二是由于FTIR对多组分的吸收谱图进行分析时, 必须借助于化学计量学。这就需要已知样品建立校正模型, 然后再对未知样品进行预测。如果校正数据与预测数据含有不同仪器的响应变量时, 则会给预测带来较大的误差。

针对上述问题, 本工作提出从模型传递的角度予以解决。模型传递就是将一台仪器上建立的校正模型, 经过处理后能够对其他相同仪器的测量信号成功预测^[5, 6]。它主要用来解决棘手样品的反复测量问题以及仪器间差异, 从而使分析过程简化、便捷。由于模型传递的主要思想是消除仪器间的差异, 而RS-FTIR的上述两个问题, 本文将其理解为样品在实测中相对于标准红外谱图出现偏差。基于此, 我们将模型传递这一思想用于RS-FTIR数据的分析, 用偏最小二乘法(PLS)对甲醇、苯、三氯甲烷和丙酮的标准红外数据和

RS-FTIR数据, 分别建立校正和预测模型, 通过正交信号处理技术(OSC), 实现了对上述混合物的RS-FTIR数据的预测。这对RS-FTIR在环境监测中的实际应用具有重要意义。

1 正交信号处理的原理

OSC是以PLS为基础建立起来的, 它以矩阵 \mathbf{X} 对矩阵 \mathbf{Y} 的信息正交为判断标准, 对 \mathbf{X} 矩阵中的变量进行过滤, 以除去 \mathbf{X} 中与 \mathbf{Y} 无关的部分^[6, 7]。其实施过程为: 首先通过PLS的分解算法计算 \mathbf{X} 的第一主成分 t , 使之与 \mathbf{Y} 正交, $t_{\text{orth}} = t - y(y^T y)^{-1} y^T t$, 然后求出权重矢量 $w_{\text{orth}} = \mathbf{X}^{-1} t_{\text{orth}}$ 和载荷矢量 $p_{\text{orth}} = \mathbf{X}^T t_{\text{orth}} / (t_{\text{orth}}^T t_{\text{orth}})$, 继而得到残差 $\mathbf{E} = \mathbf{X} - t_{\text{orth}} p_{\text{orth}}^T$, 此时相当于把第一主成分中与 \mathbf{Y} 不相关的信息去除。继续进行第二主成分计算, 用 \mathbf{E} 代替 \mathbf{X} 进行计算, 反复进行, 直至确认的 \mathbf{X} 中与 \mathbf{Y} 不相关的信息被移除。

2 实验部分

2.1 仪器与红外谱图的收集

本工作采用的是Bruker EQUINOX-55 RS-FTIR光谱仪, 分辨率为 1 cm^{-1} 。采集单组分丙酮、氯仿、甲醇和苯气体的遥感红外谱图, 每种气体收集8个浓度水平, 用积分法确定出气体浓度。然后合成出混合物样品的红外光谱图。标

收稿日期: 2005-04-18, 修订日期: 2005-08-08

基金项目: 国家自然科学基金(20175008)和南京理工大学青年学者基金(Njust200303)资助项目

作者简介: 张琳, 女, 1976年生, 南京理工大学化工学院博士研究生 * 通讯联系人

准红外谱图由 EPA 提供。图 1 给出了同样组成下标准和遥感红外谱图。其中三氯甲烷、丙酮、苯和甲醇的浓度分别为： $0.254, 0.115, 0.166, 0.068 \text{ mg} \cdot \text{L}^{-1}$ 。所有计算均在 Pentium IV 电脑上完成。

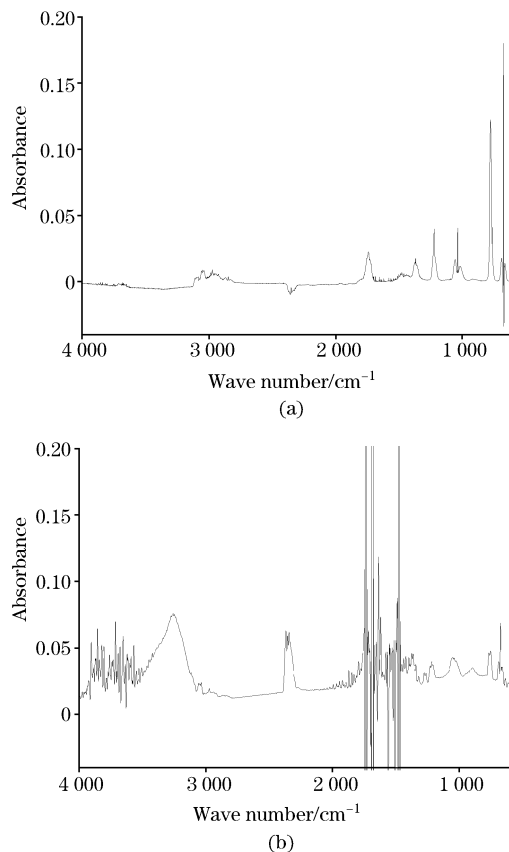


Fig. 1 The FTIR spectra of mixture from EPA and field-measurement

(a): The FTIR spectra of the mixture from EPA;

(b): The FTIR spectrum of the mixture from field-measurement

2.2 算法

文中利用 PLS 对混合物浓度进行分析,用交互验证(Cross-validation)法确定主成分数。OSC 采用 Wise 算法。用于分析的光谱区域为 $1\ 600 \sim 600 \text{ cm}^{-1}$ 。用预测样本的均方根误差(RMSEP, root mean squared error of prediction)作为评价各种方法有效性的指标。

3 结果与讨论

3.1 支集的选择

在进行 OSC 处理时,要首先确定支集。支集就是从校正集和预测集中,选取一定的样品^[5],用来进行信号处理。它必须能够包含足够的信息,来反应校正和预测数据间的差别。通常由 Leverage 算法^[5]或 Kennard-Stone 算法^[8]来确定支集。前者是根据模型逆的计算来进行选择,强调选取对校正模型有较大影响的样品,属于有标方法;后者是根据样品间相对的 Euclidean 距离来计算,以保证在数据空间的均匀

取样,属于无标方法。经优化,本工作中支集大小为 3。表 1 以三氯甲烷为例,对两种确定支集的方法进行了比较。可以看出, Kennard-Stone 选出第 4, 6, 8 个样品作为转换的标样,得到三氯甲烷的 RMSEP 为 0.002 8。而 Leverage 选出第 7, 1, 8 个样品,对应的 RMSEP 为 0.015 4。这说明在本工作中,由 Kennard-Stone 所选择的支集能够反映出实际测试样品的信号变换,因此得到的结果远优于 Leverage 算法。故本文以 Kennard-Stone 算法来选择支集。

Table 1 The effect of subset

支集选择方法	选择的样品	RMSEP
Kennard-Stone	4, 6, 8	0.002 8
Leverage	7, 1, 8	0.015 4

3.2 OSC 的实施过程

从理论部分知道, OSC 是以迭代循环过程实现的。对循环中止的确认,有不同的判断依据。文献[9]认为不能超过两次,文献[10]认为应从每个潜变量所描述变量比例的变化来判断。从 OSC 的基本原理来看,我们认为循环次数过小时, OSC 处理仅类似于基线校正;循环次数过大时, X 中已不包含 Y 的信息,因此经过 OSC 处理的 PLS 已趋近于多元线性回归(MLR)。而 MLR 的缺陷之一就是模型容易过拟合,致使预测结果不理想。在本文中,我们依次考察了不同循环次数下得到的预测结果,如表 2 所示。可见,循环次数为 5 时,计算出现非数不能进行。循环次数为 3 时,各个组分的 RMSEP 值最小。因此,在本研究体系中,循环次数确定为 3。

Table 2 The effect of number of OSC factor

循环次数	RMSEP			
	丙酮	甲醇	苯	三氯甲烷
1	0.221 5	0.034 9	0.051 9	0.013 8
2	0.217 7	0.024 9	0.068 0	0.008 3
3	0.008 5	0.018 0	0.064 0	0.002 8
4	0.079 1	0.050 2	0.091 3	0.006 7
5	—	—	—	—

“—”表示计算不能进行

3.3 预测结果

表 3 对比了采用原始数据和 OSC 校正两种方法对混合物分析得到的预测结果。可以看到,用原始数据即 EPA 标准红外谱图,建立校正模型,直接对遥感数据进行预测,得到的预测误差很大。这从图 1 中也可以看出,图 1(b)相对于图 1(a),有明显的基线漂移,同时两种方法得到的测量信号也不一致。所以由图 1(a)建模直接预测图 1(b)的数据出现了较大的误差。而经过 OSC 处理后,对丙酮、苯和三氯甲烷预测的 RMSEP 值均明显降低,但是对甲醇的预测结果改善幅度不大。图 2 以三氯甲烷为例,进一步给出了两种方法得到的预测值和真实值的对比图。这些结果都说明了 OSC 可以明显改善对 RS-FTIR 数据的预测精度。

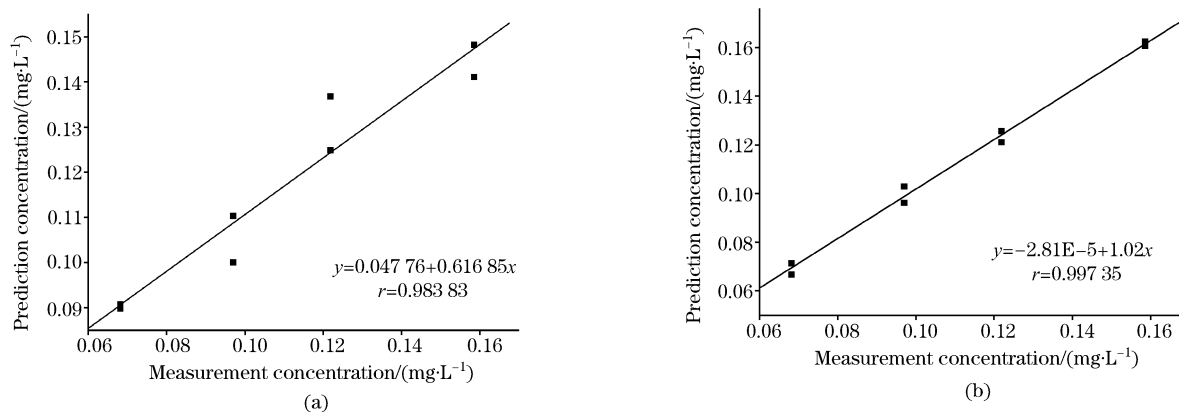


Fig. 2 Measured concentration vs. predicted concentration of chloroform with OSC

(a): Raw data; (b): With OSC

Table 3 The performance of OSC

Method	丙酮	甲醇	苯	三氯甲烷
Raw data	0.085 6	0.047 9	1.065 3	0.014 2
OSC	0.008 5	0.018 0	0.064 0	0.002 8

4 结 论

本文从模型传递的角度,建立了一种解析 RS-FTIR 谱

图的方法。对 4 种气体混合物的标准红外谱图和 RS-FTIR 数据所建立的 PLS 模型,用正交信号校正的方法进行处理,确定合适的支集选择方法以及 OSC 在实现过程中的循环次数,实现了利用标准红外数据对 RS-FTIR 数据的预测。对丙酮、甲醇、苯和三氯甲烷的预测 RMSEP 值分别为: 0.008 5, 0.018 0, 0.064 0 和 0.002 8, 而未经 OSC 处理时的 RMSEP 依次是 0.085 6, 0.047 9, 1.065 3 和 0.014 2, 其中,对丙酮、苯和三氯甲烷的预测精度有了明显改善。

参 考 文 献

- [1] Idwasi P O, Small G W, Combs R J. Appl. Spectrosc., 2001, 55(11): 1544.
- [2] LI Yan, WANG Jun-de(李 燕, 王俊德). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(6): 1104.
- [3] Frederick W Koehler IV, Small G W. Anal. Chem., 2000, 72(7): 1690.
- [4] Andersson A Claus. Chemom. Intell. Lab. Syst., 1999, 47(1): 51.
- [5] Wang Yongdong, Veltkamp D J, Kowalski B R. Anal. Chem., 1991, 63(23): 2750.
- [6] Woody N A, Feudale R N, Myles A J, et al. Anal. Chem., 2004, 76(9): 2595.
- [7] Wold S, Antti H, Lindgren Fredrik, et al. Chemom. Intell. Lab. Syst., 1998, 44(1-2): 175.
- [8] Sales F, Rius A, Callao P M, et al. Talanta, 2000, 52(2): 329.
- [9] Blanco M, Coello J, Montoliu I, et al. Anal. Chim. Acta, 2001, 434(1): 125.
- [10] Sjoblom J, Svensson O, Josefson M, et al. Chemom. Intell. Lab. Syst., 1998, 44(1-2): 229.

Calibration Transfer Used for the Interpretation of RS-FTIR Spectrum

ZHANG Lin, ZHANG Li-ming, LI Yan*, HU Lan-ping, WANG Jun-de

Laboratory of Advanced Spectroscopy, Nanjing University of Science and Technology, Nanjing 210014, China

Abstract A technique of analysis for RS-FTIR (Remote Sensing Fourier Transform Infrared) spectrum was built based on the principle of calibration transfer. A system of four-component gas mixtures, whose FTIR spectra were overlapped seriously, was determined as the object in the present research. Its FTIR data from EPA and field measurement were used to set up calibration and prediction models, respectively. With OSC (orthogonal signal correction), the model built with data of EPA predicted values of field measurement. The RMSEP values of acetone, methanol, benzene and chloroform were 0.008 5, 0.018 0, 0.064 0 and 0.002 8, respectively, while without OSC correction the corresponding values were 0.085 6, 0.047 9, 1.065 3 and 0.014 2, respectively. With optimization, the Kennard-Stone method was used for subset selection, and the iteration in the process of OSC was carried out 3 times, yielding the best performance. The result indicates that this technique can overcome the limitation of background and calibration model, which restricts the application of RS-FTIR in air monitoring.

Keywords Orthogonal signal correction; Calibration transfer; Spectrum interpretation

(Received Apr. 18, 2005; accepted Aug. 8, 2005)

* Corresponding author

《分析化学》(2007 年)

邮发代号 12-6

本刊承办广告业务

《分析化学》(ISSN 0253-3820, CODEN FHHHDT, CN 22-1125/O6)是中国科学院和中国化学会共同主办的专业性学术期刊,主要报道我国分析化学创新性研究成果,反映国内外分析化学学科前沿和进展。刊物设有研究报告、研究简报、评述与进展、仪器装置与实验技术、来稿摘登等栏目。读者对象为从事分析化学研究和测试的科技人员及大专院校师生。本刊也是有关图书、情报等部门必不可少的信息来源。

《分析化学》目前是我国自然科学核心期刊及全国优秀科技期刊,1999 年荣获首届国家期刊奖,2000 年获中国科学院优秀期刊特别奖,2001 年入选“中国期刊方阵”高知名度、高学术水平的“双高”期刊,2002 年又荣获第二届国家期刊奖和第三届中国科协优秀科技期刊奖。论文已被包括美、英、日、俄的国内外近 20 种刊物和检索系统收录。根据中国科技信息研究所历年来发布的“中国科技期刊引证报告”获悉,本刊总被引频次和影响因子均居中国科技期刊排序前列,1999 年和 2000 年其被引频次和影响因子均居化学类期刊第一名。多年来,本刊逐年被选入美国权威文摘《化学文摘》(CA)摘引量最大的 1000 种期刊(简称“CA 千种表”)中,并居我国入选“CA 千种表”期刊的前列。从 1999 年第 27 卷第一期开始被美国科学信息研究所(Institute for Scientific Information)正式收入《科学引文索引扩大版》(Science Citation Index-Expanded, SCIE, also known as SciSearch),同时还被收入《Research Alert》和《Chemistry Citation Index》等 ISI 系列。本刊是我国发行量、报道容量和国内外影响较大的科技学术期刊之一。

本刊为月刊,2005 年每期从 128 页增至 144 页(大 16 开),由科学出版社出版。国内单价 15.00 元,全年 180.00 元。邮发代号 12-6,全国各地邮局订阅,国外代号 M336,中国国际书店订购,漏订读者,可与编辑部联系。广告经营许可证号:2200004000094,广告代理:北京行胜言广告有限公司,电话:010-84411842。

编辑部地址:长春市人民大街 5625 号 邮政编码:130022

编辑部电话:(0431)5262017/5262018 传真:(0431)5262018

E-mail: fxxh@ciac.jl.cn

网址: <http://www.analchem.cn>